A Description and Interpretation of the NPGS Birdsfoot Trefoil Core Subset Collection

J. J. Steiner,* P. R. Beuselinck, S. L. Greene, J. A Kamm, J. H. Kirkbride, and C. A. Roberts

ABSTRACT

Systematic evaluations for a range of traits from accessions in National Plant Germplasm System (NPGS) core subset collections should assist collection management and enhance germplasm utilization. The objectives of this research were to: (i) evaluate the 48-accession NPGS birdsfoot trefoil (*Lotus corniculatus* L.) core subset collection by means of a variety of biochemical, morphological, and agronomic characters; (ii) determine how these characters were distributed among the core subset accessions and associations among plant and ecogeographic characters; (iii) define genetic diversity pools on the basis of descriptor interpretive groups; and (iv) develop a method to utilize the core subset as a reference collection to evaluate newly acquired accessions for their similarity or novelty. Geographic information system (GIS) databases were used to estimate the ecogeography of accession origins. Interpretive groups were constructed to describe the range of core subset descriptor variation by cluster analysis and verified by discriminant analysis. Associations among plant descriptors and with ecogeographic characteristics were determined by Pearson's correlation coefficients or the Mantel *Z* statistic. The accessions were classified into four distinct genetic diversity pools that were described by plant traits and ecogeographic origins. The core subset used as a reference collection successfully classified three unique accessions not originally included in the core subset. This approach identified germplasm that was different from that present in most North American cultivars and can be used to evaluate future acquisitions. The concepts of interpretive groups, genetic diversity pools, and reference collection comparisons should be applicable for assessing and managing other core subset collections.

The large size of some germplasm collections can hinder their effective utilization (Frankel, 1984), so the core subset concept was proposed as a way to facilitate germplasm use in crop improvement (Brown, 1989). However, the lack of available characterization information about accessions in ex situ germplasm collections is a common problem that can constrain the utilization of collections (Marshall, 1989). Describing the range of genetic variation in core subsets should provide users valuable information about individual accessions, relationship among traits, and the structure of collections (Beuselinck and Steiner, 1992). With such information, core subsets could be examined for specific traits of interest in lieu of having to evaluate all accessions in the active collection. This should increase evaluation efficiency, reduce the number of accession requests, and reduce the frequency and costs for collection regeneration (Greene and McFerson, 1994). Well characterized collections can help plant explorers develop plans for collecting unique genetic materials that are not represented by present ex situ collection holdings (Steiner and Greene, 1996; Greene et al., 1999a,b).

In the past, most core subset collections were based on just a few traits or the geographic origins of the accessions. Assessing a range of collection descriptors, including plant morphology, biochemistry, and collecting site habitat, has advantages that may be lost when only a single trait is considered. Differences among accessions based on molecular markers may indicate genome-wide levels of genetic variation (Avise, 1994). However, molecular markers variation may frequently be neutral (Kimura, 1982), and morphologic traits can converge when exposed to similar selection pressures (Rieseberg and Brunsfeld, 1992; Steiner and Garcia de los Santos, 2001). Also, when one or just a few traits are used to characterize collections, there is less opportunity to understand relationships among different traits within collections (Brown, 1989). Multivariate methods, including cluster and principal components analyses, have been used to create descriptive profiles and classify plant characters (Spagnoletti Zeuli and Qualset, 1987; Warburton and Smith, 1993). Examining how accessions and traits are distributed among classification groups can also provide insights into collection genetic diversity and identify sources of genetic variation that may not be currently utilized.

The objectives of this research were to: (i) evaluate the 48-accession NPGS birdsfoot trefoil core subset collection by means of a variety of biochemical, morphological, and agronomic characters; (ii) determine how these characters were distributed among the core subset accessions and associations among plant and ecogeographic characters; (iii) define genetic diversity pools based on descriptor interpretive groups; and (iv) develop a method to utilize the core subset as a reference collection to evaluate newly acquired accessions for their similarity or novelty

MATERIALS AND METHODS

The 48 accessions included in the birdsfoot trefoil core subset collection were chosen in 1981 to include the maximum number of countries of origin among the accessions available from the USDA-ARS NPGS base collection (P. Beuselinck,

J.J. Steiner and J.A. Kamm, National Forage Seed Production Res. Cntr., USDA-ARS, 3450 SW Campus Way, Corvallis, OR 97331; P.R. Beuselinck, USDA-ARS, Plant Genetics Research Unit, Columbia, MO 65211; S.L. Greene, National Temperate Forage Legume Germplasm Resources Unit, Prosser, WA 99350; J.H. Kirkbride, Systematic Botany and Mycology Lab., Beltsville, MD 20705; and C.A. Roberts, Dep. of Agronomy, Univ. of Missouri, Columbia, MO 65211. Oregon Agric. Exp. Stn., Technical Paper No. 10,968. The use of trade names in this publication does not imply endorsement of the products named or criticism of similar ones not mentioned. Received 15 Dec. 2000. *Corresponding author (steinerj@ucs.orst.edu).

Published in Crop Sci. 41:1968–1980 (2001).

Abbreviations: FRI, flowering response index; NPGS, National Plant Germplasm System; Pool, genetic diversity pool; PC, principle components; RAPD, random amplified polymorphic DNA; r = Pearson's correlation coefficient; r_{pm} = Mantel Z product moment correlation; and SGPP, high salt soluble seed globulin polypeptides.

1992, personal communication). At least one accession from each of the 21 countries represented in the active collection was included in the subset. The core subset included 12 cultivars, two germplasm releases, 27 wild or naturalized populations collected in the Old World, and seven accessions whose Old World ecogeographic origin were unknown. All core subset accessions are currently available from the 363 accession NPGS collection.

Plant Descriptors

Random Amplified Polymorphic DNA (RAPD)

Fourteen plants from each accession were grown from seeds in the greenhouse in 1992. Three young leaves from each plant were collected, bulked, frozen over liquid N_2, and lyophilized. Dried leaf material was ground into a fine powder and stored at room temperature. DNA was extracted from two replicates of 30 to 40 mg of ground material from each bulked sample as described in Steiner and Garcia de los Santos (2001). The extracted DNA samples were diluted to approximately 1 ng L^{-1} for use in the RAPD reactions.

Approximately 5 ng of DNA was used as a template for the RAPD reactions in a 25-μL mixture containing 50 m*M* Tris-HCl pH 9, 45 m*M* ammonium sulfate, 1.5 m*M* magnesium chloride, 100 μ*M* dNTPs, 0.2 μ*M* 10-mer primer, 1 U *Tfl* DNA polymerase, and overlaid with 50 μL mineral oil. The primers used for screening the accessions were OPA-8, OPA-10,OPB-6, and OPB-13 (Operon, Alameda, CA). The reactions were conducted in a MJ Research (Watertown, MA) 96-well microtiter format thermocycler (PTC-100) using the temperature profile: 95°C for 3 min, 46°C for 1 min, 72°C for 1 min, 41 cycles of 1 min each at 94, 46, and 72°C, and a final 9 min at 72°C. The RAPD products were separated by electrophoresing 11 μL of the reaction mixture in a 17.5 g L^{-1} 3:1 NuSieve agarose gel (Cambrex Corporation, East Rutherford, NJ) in 1× Tris-borate EDTA. The gels were run at a constant 90 mA with a 100-bp DNA ladder standard included in each gel. The bands were visualized, photographed, scanned, and scored as described by Steiner and Garcia de los Santos (2001). A total of 37 RAPD bands were selected for scoring.

Seed Globulin Polypeptides (SGPP)

The SGPP descriptors for the core subset were obtained from a 128 accession data set that reported 13 high salt-soluble polypeptides ranging in weight from 23.1 to 65.3 kDa. A complete report of the methods and materials were documented in Steiner and Poklemba (1994).

Herbage Tannin Content

Herbage tannin content (tannin) for all accessions was measured from plants grown in the greenhouse at Columbia, MO in 1994. Herbage was lyophilized and tannin content determined by near infrared reflectance spectroscopy. The methods and materials used were documented in Roberts et al. (1993).

Flowering

Seedlings of each accession grown in the greenhouse were transplanted 0.5 m apart in rows 1 m apart in randomized complete block designs with five replications of 20 plants each during the fall of 1989 at Corvallis, OR. Field flowering intensity was scored from 1 to 5: 1, not flowering; 2, very few flowers; 3, few to moderate flowering; 4, moderate flowering; and 5, intense flowering. Measurements at Corvallis were made five times during the flowering period in 1990 (7 and 20 May, 1 and 20 June, and 20 September) and three times in 1991 (4 May, 5 June, and 18 September). The percentage of plants of each accession that had initiated flowers by 7 May 1990 at Corvallis was determined. The effects of 13-, 16-, and 19-h photoperiod lengths on flowering was obtained from a 68 accession data set (Steiner, 2000, unpublished data). The flowering response index (FRI) was calculated as:

$$\mathrm{FRI} = \left(\frac{\sum_{i=1}^{t} \frac{n_i}{T_i}}{\sum_{i=1}^{t} \frac{n_i}{N}} \right) / \mathrm{HU}_{\max} \times 10^6 \qquad [1]$$

where n_i is the number of clones that flowered at Time i, T_i is the number of heat units at Time i that a clone was observed to flower, N is the number of clones for an accession that were used, HU_{max} is the heat unit cut off threshold for observations at all photoperiods, and t is the number of counts made until HU_{max} was reached for an accession. For this experiment, HU_{max} was 1550, determined on the basis of 23°C light and 15°C dark period temperatures and a base temperature of 10°C.

Seed Chalcid Susceptibility and Plant Morphology

Seedlings of each accession grown in the greenhouse were transplanted 0.5 m apart in rows 1 m apart in randomized complete block designs with three replications during spring of 1990 at Columbia, MO. Approximately 20 mature, unshattered, umbels were collected from each accession in each of three replicates in late-July 1991. The umbels were placed in one liter hard-paper containers with lids (ice cream containers) and put in a growth chamber with ventilation at 24°C for 6 wk to promote chalcid (*Bruchophagus platypterus* Walker) emergence. The average number of chalcids per umbel per replicate was determined by counting and removing the emerged chalcids. The dried contents of each container were weighed to estimate pod weight minus chalcid weight.

Leaflet length and width of the central leaflet of the fourth leaf axil from the distal end of five stems of three plants at least nine months old were multiplied as an estimate of leaflet area at Corvallis in 1992. The leaflet indumentum were classified as glabrose, pilose, or pubescent. General plant growth habit (ascending and prostrate) and plant color (typical dark green and light green) were also recorded. Accession descriptions of plant morphology are summarized in Steiner and Poklemba (1994).

Cytology and Taxonomy

All accessions had somatic chromosome numbers of $2n = 4x = 24$ as determined by root tip squashes Beuselinck et al., 1996). The chromosomes were not karyotyped to determine whether other members of the *L. corniculatus* group that closely resemble birdsfoot trefoil were inadvertently included in the sample (Larsen, 1954; Small et al., 1984). Pressed plant specimens were examined to verify that all accessions were *L. corniculatus* (J. Kirkbride, 1995, personal communication).

Ecogeographic Origin

The origins of the core subset accessions were estimated from collection site data reported in the USDA-ARS Germplasm Resources Information Network (GRIN).The latitude and longitude of the accession collecting sites were determined from original passport information or estimated by retroclassification when actual collecting site coordinates were not recorded (Steiner and Greene, 1996). The ecogeographic data

describing the collecting sites were acquired from the U.S. Environmental Protection Agency and National Oceanic and Atmospheric Administration (EPA/NOAA) Global Ecosystems Databases (Kineman and Ohrenschall, 1992, 1994). The 48 collecting sites were described by the following: lowest (low) and highest (high) monthly temperature, monthly accumulated precipitation (precipitation) and monthly percentage of sunshine hours (sunshine) (Leemans and Cramer, 1992), and monthly accumulated snow depth (snow) (Chang et al., 1990). Collecting sites were classified as being either of lowland or highland origin on the basis of Bailey's (1989) Ecoregions of the Continents map. The geographic distances between collecting sites were estimated by a great arc approach developed by A. Afonin, Vavilov Research Institute, St. Petersburg, Russia:

$$D_{\text{geog}} = \sqrt{\begin{aligned}&\{(\text{Long}_a - \text{Long}_b \times (\pi r/360)\\ &\times [\cos(\text{Lat}_a) + \cos(\text{Lat}_b)]\}^2 + \{(\pi r/180)\\ &\times [\cos(\text{Lat}_a) + \cos(\text{Lat}_b)]\}^2\end{aligned}} \quad [2]$$

where Long_a and Long_b are the longitudes of Collecting Sites *a* and *b* being compared, respectively; cos (Lat_a) and cos (Lat_b) are the cosin of the latitudes of collecting sites a and b being compared, respectively; and r is the 6378 km radius of the earth (Steiner and Garcia de los Santos, 2001).

Statistical Analyses

Interpretive Group and Genetic Diversity Pool Classifications

Data analysis methods were performed in specific sequence (Fig. 1). Categorical grouping classes for the nine plant descriptors used to describe the core subset were assembled by means of cluster analysis based on Euclidean distance and Ward's (1963) clustering technique (Systat for the Macintosh, Evanston, IL). These categorical classes, named *interpretive groups*, were used to classify the accessions on the basis of the multistate descriptors: (i) the combined effects of the qualitative presence or absence values for 37 polymorphic RAPD product bands; (ii) the combined effects of the standard normal deviates (Snedecor and Cochran, 1980) from the relative band intensity of the 13 SGPP bands (Steiner and Poklemba, 1994); (iii) the combined effects of the standard normal deviate of the five flower intensity observations in the field (flower pattern) measured in 1990 at Corvallis, OR; (iv) the combined effects of three photoperiod length treatments on flowering as measured by the FRI; and simple quantitative single-state descriptors (v) percentage of clones that had flowered by 7 May 1990 at Corvallis, OR; (vi) herbage tannin content; (vii) central leaflet size (width × length); (viii) seed pod weight at Columbia, MO; and (ix) seed chalcid infestation (number per umbel) at Columbia, MO. The number of interpretive group classes for each of the nine descriptors was determined by examining the cluster analysis dendrograms for each. Simple quantitative descriptors were examined by means of two to six classes in a series of analysis of variance tests. The optimal number of classes (c_{opt}) was based on the maximum *F*-statistic and separations among all class means for an interpretive group based on Fisher's least significant difference using:

$$c_{\text{opt}} = \lim [D_{\text{n}} \geq 0.5 \cdot D_{\text{g}}]; \text{ whenever } n > 2 \quad [3]$$

where D_{g} was the greatest amalgamation distance between two clusters and D_{n} is the least successive amalgamation distance between two clusters and is greater than or equal to one-half D_{g}.

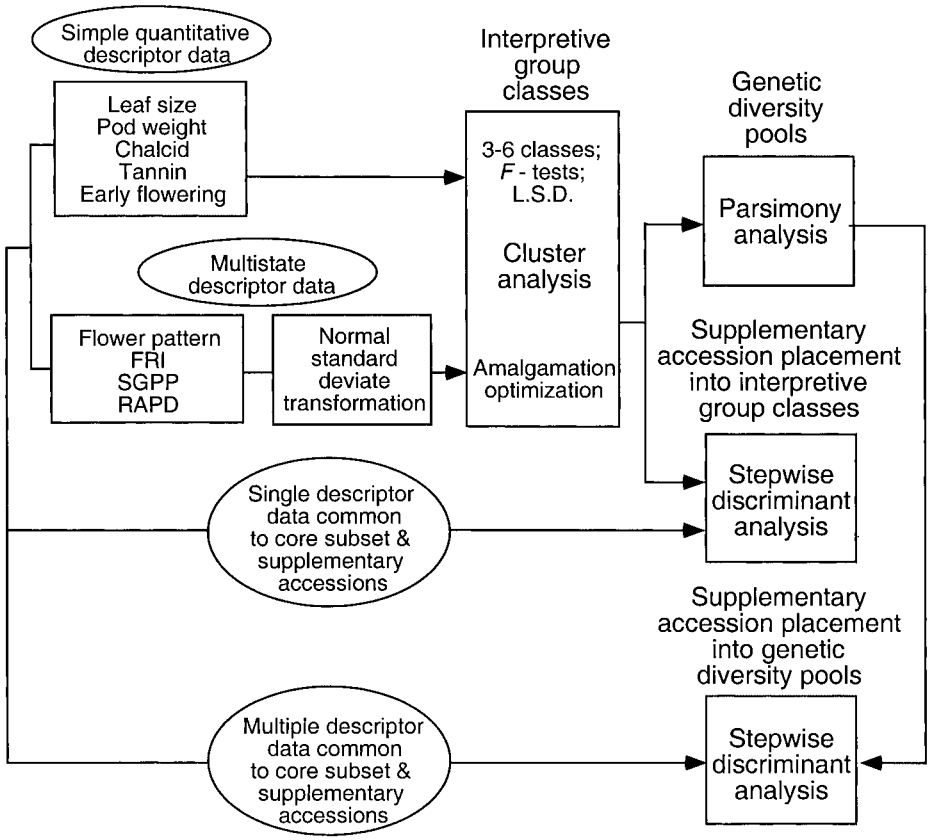

Fig. 1. Data analysis flow chart for interpreting the birdsfoot trefoil core subset collection.

By means of PAUP 4.0 (Swofford, 1998), the class data from the nine interpretive group plant descriptors were used to classify the core subset accessions into genetic diversity pools (Fig. 2). The options selected for the analysis were heuristic search, maximum trees equals 1000, 50% majority rule for the unrooted consensus tree, and group frequencies for the placement of taxa in branches. Each accession was assigned to one of four groups on the basis of a visual examination of the tree to determine likely groupings of accessions by similar geographic origin. The groupings of accessions were designated *genetic diversity pools.* The significance of each interpretive group descriptor in developing the four genetic diversity pools and the percentage of correctly classified cases were tested by Wilks' Lambda statistic by step-wise discriminant analysis (SPSS Inc., Chicago, IL) with the parsimony analysis results used as the grouping variable. The percentage of accessions from the core subset that were correctly classified was determined. Misplaced accessions were reassigned to highest probability groups and the analyses rerun to determine whether the analysis results changed. The maximum number of significant canonical discriminant functions needed to describe the core subset was also determined. After the accessions were placed into a genetic diversity pool, the mode for each interpretive group descriptor was determined. A similarity index was calculated for each descriptor within each genetic diversity pool by:

$$I = \left\{S - \left[\sum_{k \in n} diff(x_{ik}, x_{jk}, \ldots x_{nk})\right]\right\}/S \qquad [4]$$

where the similarity index (I) for a set of observations with S possible comparisons for an interpretive group descriptor (k), with a genetic diversity pool with n accessions, and x_{ik}, x_{jk}, $\ldots$ x_{nk} being the k states of the descriptor in the genetic diversity pool. The S possible combinations of interpretive group states for comparison in a genetic diversity pool were determined by:

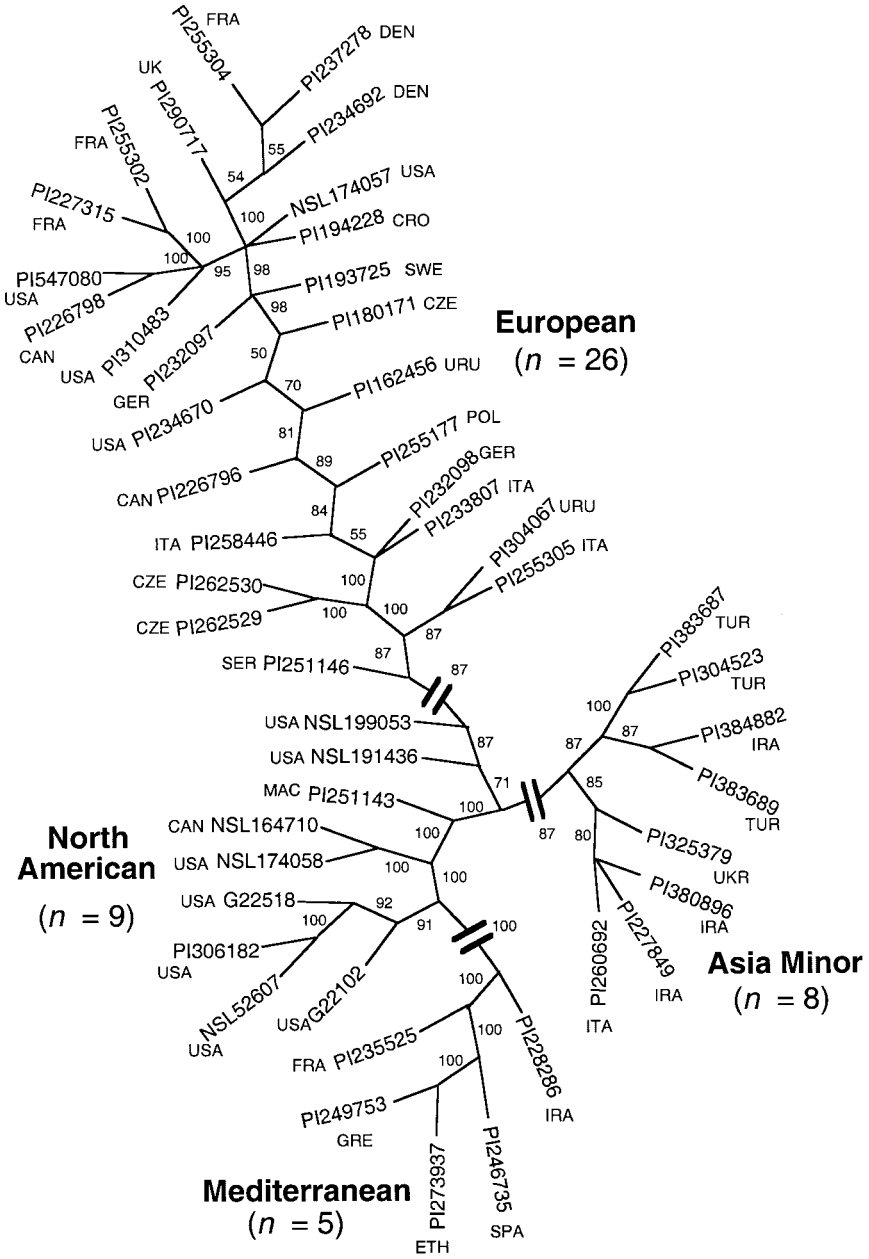

Fig. 2. Four birdsfoot trefoil core subset genetic diversity pools based on nine interpretive group plant descriptors by parsimony analysis.

$$S = \left[\frac{n!}{(n - r)!}\right]/2 \quad [5]$$

where n is the total number of accessions in a genetic diversity pool and $r = 2$ (two accessions compared at a time). When all x_{ik} states of k in a genetic diversity pool are the same, $I = 1.0$.

Three accessions (G 31276 from Morocco, PI 234670 from France, and PI 234811 from Switzerland), were not included in the core subset but were of interest because of unique morphology and were designated as supplementary accessions to the core subset. The core subset was used as a reference collection (Beuselinck and Steiner, 1992) to predict the placement of the three supplementary accessions into core subset genetic diversity pools. The classification was based on the simple quantitative (leaf, tannin, and FRI) and multistate (SGPP and RAPD) descriptor data common to the core subset and supplementary accessions.

Plant and Ecogeographic Descriptor Relationships

Pearson's correlation coefficient (r) was used to describe the associations among the plant descriptors with one another (chalcid, pod, and tannin) and with collecting site ecogeographic descriptors (low and high temperature, sunshine, and precipitation). A subset of 27 wild accessions was analyzed for associations among plant descriptors and plant with ecogeographic descriptors as above to determine the impact of nonwild accessions on correlation coefficients describing the entire core subset.

Congruence among the multistate plant descriptors (RAPD, SGPP, and flowering pattern) was determined (Steiner and Garcia de los Santos, 2001) by means of the MXCOMP command in NTSYSpc version 2.2 (Rohlf, 1997) to calculate the product moment correlations (r_{pm}) derived from the normalized Mantel Z (Mantel, 1967). The associations for the multistate descriptors with early flowering, tannin, pod, chalcid, leaf, and FRI were also determined. Symmetric Euclidean distance matrices were calculated from the following:

$$A = \begin{bmatrix} a_{011} & a_{012} & \dots & a_{01j} \\ a_{021} & a_{022} & \dots & a_{02j} \\ \vdots & \vdots & \vdots & \vdots \\ a_{281} & a_{282} & \dots & a_{28j} \end{bmatrix}$$

where A was any $i \times j$ matrix with i equaling one to 48 of the core subset accessions and j was the number of levels measured for the multistate descriptive variable. The Euclidean distance matrices were constructed using Systat 5.2.1 for the Macintosh (Evanston, IL). All levels of a descriptive variable were entered into a single line for each of the 48 accessions and transposed. The STATS/CORR/EUCLIDIAN function was used to produce the following distance matrix (D):

$$A = \begin{bmatrix} e_{0101} & e_{0102} & \dots & e_{0128} \\ & e_{0202} & \dots & e_{0228} \\ & & \dots & e_{ij} \\ & & & e_{2828} \end{bmatrix}$$

where e_{ij} off-of-the-diagonal was the Euclidean distance for each of the 48 core subset accessions with the other 47 genotypes. Plots of one distance matrix against the other were determined with the values along the diagonal ignored. In performing the normalized Mantel Z test, if both matrices being compared contained corresponding distance estimates, then the value of the Z test criterion was large compared with chance expectation, and was determined by:

$$Z = \sum_{i<j}^{n} x_{ij} y_{ij} \quad [6]$$

where X_{ij} and Y_{ij} are two different measures relating to ith element of a sample to the jth element in both matrices. The estimated Z value was compared with its permutational distribution obtained from 500 random samples of all possible permutations of the matrices and was the measure of the degree of closeness between the two matrices.

RESULTS AND DISCUSSION

Core Subset Classification

The NPGS birdsfoot trefoil germplasm collection contained 335 accessions at the time this study was initiated. Forty-eight accessions acquired before 1981 were selected in 1992 from the active collection on the basis of country of origin and were designated as the core subset (P.R. Beuselinck, 1992, personal communication) and represented 14% of the total number of accessions (Table 1). Most of the birdsfoot trefoil accessions acquired before 1981 had relatively little passport collecting site information that could be used to describe their exact ecogeographic origins (S.L. Greene, 1992, personal communication). Since a strict sampling regime for selecting the core subset accessions was not used, the frequency of occurrence of accessions by origin and the different classes shown here should not be used to infer relative distribution frequency in the NPGS active collection.

Few core subset accessions had been characterized for a wide array of characters to determine sufficiently the range of genetic variation available in the collection. Four previous characterizations of accessions from the NPGS birdsfoot trefoil active collection have been published: agronomic and herbage quality measurements and in vitro digestible dry matter (McGraw et al., 1989); high salt soluble seed globulin polypeptides (Steiner and Poklemba, 1994); susceptibility to foliar and shoot blight caused by *Rhizoctonia* spp. (English and Beuselinck, 2000); and a genetic, morphological, ecological, and crossing ability assessment of wild genotypes (Steiner and Garcia de los Santos, 2001; Garcia de los Santos et al., 2001).

Interpretive Groups

We proposed and used interpretive groups derived from cluster analyses of both simple and multistate characters (nine descriptors total) to classify each birdsfoot trefoil core subset accession by the nine descriptors. Each interpretive group was comprised of three to five categorical classes per descriptor (Table 1). Both the range of variation for simple quantitative descriptors (e.g., seed chalcid number and herbage tannin content, Table 2) and summaries of the multistate character descriptors (FRI, Table 2; SGPP, Table 3; RAPD, Table 4; flowering intensity rating pattern, Fig. 3) were determined.

Interpretive groups have the benefit of allowing germplasm data sets to be sorted by various descriptors to identify individual or groups of accessions that contain

Table 1. Categorization of nine descriptors as interpretive groups and combined genetic diversity pools describing the 48-accession birdsfoot trefoil core subset from the USDA-ARS National Plant Germplasm System collection and placement of three supplementary accessions into genetic diversity pools.

			Interpretive group descriptor†								
Entry‡	Cultivar	Origin	Leaf size (5)§	Pod weight (5)	Chalcid (4)	Tannin (3)	Early flowering (4)	Flower pattern (4)	FRI (3)	SGPP (4)	RAPD (4)
						Genetic Diversity Pool Asia Minor					
PI 227849w		Iran	1	2	3	1	1	2	3	1	3
PI 260692w		Italy	1	2	2	1	4	3	3	1	3
PI 304523w		Turkey	4	1	2	1	1	2	2	3	3
PI 325379w		Ukraine	1	2	2	1	1	2	1	1	3
PI 380896w		Iran	1	2	2	1	1	2	3	1	4
PI 383687w		Turkey	1	1	1	1	1	2	3	3	3
PI 383689w		Turkey	3	2	3	1	1	2	2	3	1
PI 384882w		Iran	1	2	3	1	1	2	2	3	3
		Interpretive group mode:	1	2	2	1	1	2	3	1, 3	3
		Similarity index#:	0.54	0.57	0.32	1.00	0.75	0.75	0.32	0.43	0.46
						European					
NSL 174057	NC-83	USA	2	2	3	2	1	1	2	1	2
PI 162456		Uruguay	2	3	4	2	2	3	2	3	2
PI 180171w		Czech Republic	2	5	4	2	2	2	2	2	2
PI 193725w		Sweden	2	5	3	2	2	3	2	1	2
PI 194228w		Croatia	2	2	3	2	2	3	3	1	2
PI 226796		Canada	2	3	4	2	3	3	2	1	2
PI 226798		Canada	3	3	3	2	2	3	2	1	2
PI 227315w		France	3	2	4	2	2	3	2	1	2
PI 232097		Germany	2	5	3	2	3	3	2	1	2
PI 232098		Germany	3	4	4	2	3	3	2	1	2
PI 233807		Italy	3	4	4	2	3	3	2	1	2
PI 234670	Kalo	USA	2	1	4	2	4	4	1	1	2
PI 234692w		Denmark	1	2	3	2	3	3	1	3	2
PI 237278	Late Roskilde II	Denmark	2	2	3	2	2	3	1	3	3
PI 251146w		Serbia	4	4	1	2	1	2	2	1	3
PI 255177w		Poland	3	3	4	2	3	3	3	2	2
PI 255302w		France	3	2	4	2	2	3	2	1	2
PI 255304w		France	2	2	2	2	2	3	1	3	2
PI 255305w		Italy	4	4	3	2	3	3	2	3	3
PI 258446	Stirpe 13	Italy	3	1	2	2	3	3	2	1	2
PI 262529w		Czech Republic	4	4	4	2	3	3	3	1	2
PI 262530w		Czech Republic	4	4	4	2	3	3	3	1	2
PI 290717w		United Kingdom	1	2	2	2	2	3	2	3	2
PI 304067		Uruguay	3	3	3	2	3	2	2	3	3
PI 310483	Viking	USA	3	2	3	1	2	3	2	1	2
PI 547080	Au Dewey	USA	3	3	3	2	2	3	3	1	2
		Interpretive group mode:	3	2	3	2	2	3	2	1	2
		Similarity index:	0.25	0.19	0.32	0.92	0.34	0.66	0.47	0.48	0.73
						Mediterranean					
PI 228286w		Iran	1	2	1	1	1	2	2	1	4
PI 235525w		France	5	2	1	3	1	1	1	4	4
PI 246735w		Spain	5	2	1	3	1	4	2	4	4
PI 249753w		Greece	4	3	3	3	1	4	2	1	4
PI 273937w		Ethiopia	5	4	3	3	2	4	3	4	4
		Interpretive group mode:	5	2	1	3	1	4	2	4	4
		Similarity index:	0.30	0.30	0.40	0.60	0.60	0.30	0.30	0.40	1.00
						North American					
G 22102	Fargo	USA	4	2	1	1	1	1	2	1	1
G 22518	Empire	USA	3	2	1	1	1	1	2	1	1
NSL 52607	Dawn	USA	3	2	1	1	1	1	1	1	1
NSL 164710	Cree	Canada	1	3	3	1	1	1	1	1	1
NSL 174058	Norcen	USA	1	3	4	1	1	1	2	1	1
NSL 191436	Fergus	USA	1	3	4	1	1	2	2	1	3
NSL 199053	MU-81	USA	4	5	4	1	1	2	2	2	1
PI 251143w		Macedonia	1	2	4	2	1	1	2	1	3
PI 306182	Leo	USA	3	2	1	1	1	1	1	1	1
		Interpretive group mode:	1	2	1, 4	1	1	1	2	1	1
		Similarity index:	0.28	0.36	0.33	0.78	1.00	0.61	0.50	0.81	0.61
						Supplementary Accessions to the Core Subset Collection¶					
G 31276	EU	Morocco	3	–	–	1	–	–	2	1	2
PI 234670	EU	France	3	–	–	1	–	–	1	3	3
PI 234811	AM	Switzerland	3	–	–	1	–	–	3	1	3

† Descriptors: Leaf, size of center leaflet (length × width); Pod, weight of an individual dried umbel of pods; Chalcid, number of seed chalcids per umbel; Tannin, herbage tannin content; Flower, flowering pattern for five sampling times; Early flowering, percentage of clones that had flowered on 7 May; FRI, flowering response index; RADP, random amplified polymorphic DNA; and SGPP, high salt-seed globulin polypeptides.

‡ Genetic diversity pools were determined by a maximum parsimony analysis using the multistate classes of the eight interpretive descriptor groups. The names of the genetic diversity pools are: AM, Asia Minor; EU, European; ME, Mediterranean; and NA, North American. The (w) indicates the accession is from a wild or naturalized Old World collecting site.

§ Numbers in parentheses indicates the number of classes for each interpretive descriptor group.

¶ Placement of three supplementary accessions using a stepwise discriminant analysis function based on herbage tannin content, 13 hr FRI, and eight RAPD products from the core subset accessions using a maximum $F\ P \leq 0.01$ to enter, and a minimum $F\ P \leq 0.05$ to remove each variable. EU and AM indicates placement into the European and Asia Minor genetic diversity pools, respectively.

Similarity index is the percentage of possible interpretive group class comparisons in a genetic diversity pool that are the same.

Table 2. Differences among interpretive group categories for six quantitative descriptors describing the 48 accession core subset of the USDA-ARS NPGS birdsfoot trefoil collection.

Descriptor†	Interpretive group‡	Descriptive statistic				
		n	Minimum	Maximum	Mean	SEM§
			no. umbel^{-1}			
Chalcid	1	9	0.7	2.4	1.7	0.21
	2	7	2.7	3.8	3.3	0.15
	3	17	4.3	5.8	5.2	0.11
	4	15	6.5	8.6	7.3	0.17
			m eq. g^{-1}			
Tannin	1	18	2.0	9.0	4.0	0.49
	2	26	17.0	33.0	24.6	0.85
	3	4	53.0	98.0	71.3	9.50
			mm mm^{-1}			
Leaf	1	13	12.8	26.1	18.9	1.27
	2	10	28.0	36.0	31.3	0.75
	3	14	39.6	55.8	46.0	1.62
	4	8	61.9	88.2	75.1	3.91
	5	3	117.0	144.0	133.2	8.24
			g $\times$ 10^2			
Pod weight	1	4	3.8	4.3	4.1	0.11
	2	23	4.7	6.1	5.4	0.08
	3	10	6.5	7.0	6.6	0.06
	4	7	7.2	8.1	7.7	0.10
	5	4	8.8	9.3	9.1	0.10
			%			
Early flowering	1	22	0.0	7.5	2.1	0.53
	2	13	13.3	27.5	20.6	1.31
	3	11	29.2	39.2	32.8	0.98
	4	2	60.8	69.2	65.0	4.20
FRI	1	9	2.3	3.1	2.8	0.10
	2	29	3.3	4.1	3.8	0.04
	3	10	4.2	5.8	4.8	0.17

† Chalcid, number of seed chalcids per umbel; Tannin, herbage tannin content; Leaf, size of center leaflet (length $\times$ width); Pod weight, weight of an individual dried umbel; Early flowering, percentage of clones that had flowered by 7 May at Corvallis, OR; and Flowering Response Index (FRI), the sum of the FRI for plants grown under 13, 16, and 19 hour day length periods.
‡ Interpretive groups were based on cluster analysis results using Euclidean distance and Ward's clustering technique.
§ SEM, standard error of the mean.

certain combinations of desired characters. Interpretive groups also allow comparisons of qualitative or quantitative descriptors on a common scale. Single observation descriptor data (e.g., leaf size, herbage tannin content, and growth habit) are straight forward ordinal or quantitative measures that can be easily presented as character profiles and directly interpreted. However, when more complex measures comprised of multistate character observations are used (e.g., flowering response measured as a series of temporal observations, or numerous RAPD product bands), the data need to be reduced into a more usable form that can be quickly interpreted.

Since all of the variation from multistate descriptors can be directly understood and utilized by means of interpretive groups, information about an accession is not lost through the classification process (e.g., specific band products). Interpretive groups are thus presented as an alternative to principle components analysis. Principal component analysis has traditionally been used to reduce large multiple variable data sets into more usable forms, but was developed with the intention of having the first two or three principal components account for a significant portion of the total variation measured (Goodman, 1972). A difficulty with principal component analysis is that the principle component (PC) values cannot be directly used to interpret the effects of specific variable attributes.

When principal component analysis has been used with data describing germplasm collections, species-level descriptors may sometimes account for only a small

Table 3. Interpretive group categories for 13 high-salt soluble seed globulin polypeptides describing the 48 accessions from the USDA-ARS NPGS birdsfoot trefoil core subset collection.

Interpretive group†	*n*	Number of bands‡	Polypeptide (kDa)												
			65.34§	62.14	55.66	54.16	52.80§	50.44	47.94	47.04	36.92	30.16§	29.25	23.49	23.08
1	31	5 (4,5,6,7,8)	(−)¶	(−)	−	+	−	(−)	(+)	(−)	≈	−	+	+	+
2	3	5 (6)	−	−	+	+	−	−	(−)	−	−	−	+	+	+
3	11	7 (6)	≈	−	−	+	−	(−)	+	(−)	(+)	−	+	+	+
4	3	8 (10)	(−)	+	(−)	(−)	(−)	+	(−)	(+)	(−)	+	+	+	+

† Based on cluster analysis using the standard normal deviate of the relative intensity of the detected seed polypeptides from each accession.
‡ Numbers in parentheses less than or greater than the band number given indicate band deletions or additions, respectively.
§ Indicates SGPP band used in stepwise discriminant analysis function to group accessions into genetic diversity pools.
¶ +, −, (+), (−), and ≈ indicate the presence, absence, mostly present (>60%), mostly absent (<40%), or equally absent and present bands (40−60%), respectively, of the polypeptide class indicated for each interpretive group.

Table 4. Summary of the occurrence of 37 random amplified polymorphic DNA products describing the 48-accession USDA-ARS NPGS birdsfoot trefoil core subset collection.

RAPD primer	RAPD product	Interpretive group†			
		1	2	3	4
	base pairs				
B13‡	1880	+§	+	+	≈
B13	1479	≈	≈	(−)	(−)
B13	873	(−)	(−)	≈	(−)
A10‡	1981	(−)	(−)	(−)	≈
A10	1720	(−)	(−)	(−)	≈
A10	1665	(−)	(−)	≈	−
A10‡¶	1323	−	−	≈	(+)
A10	1253	+	+	(+)	−
A10	1110	≈	≈	(+)	−
A10	971	(+)	(+)	(+)	(+)
A10	689	(+)	(+)	(−)	(−)
A10	672	≈	≈	(+)	(−)
A10	618	+	+	(+)	(−)
A10	560	≈	≈	(−)	−
A10	530	≈	≈	≈	(−)
A10	332	(−)	(−)	(−)	≈
B6‡¶	1332	(+)	(+)	(−)	(−)
B6	1315	≈	≈	(−)	+
B6¶	1306	(−)	(−)	(+)	−
B6	1256	(−)	(−)	≈	(−)
B6	1070	≈	≈	(−)	(−)
B6	1011	≈	≈	(−)	−
B6	986	≈	≈	(+)	(−)
B6	938	≈	≈	(−)	(−)
B6¶	780	(−)	(−)	(−)	−
B6	634	+	+	(+)	−
B6‡¶	466	+	+	+	(+)
A8	1773	(+)	(+)	≈	≈
A8	1393	(−)	(−)	≈	(−)
A8	1349	(−)	(−)	≈	≈
A8	1072	(−)	(−)	−	(+)
A8	1027	(−)	(−)	(−)	(−)
A8	969	+	+	(+)	−
A8‡	858	(−)	(−)	(−)	−
A8	821	(+)	(+)	≈	(+)
A8	781	(−)	(−)	≈	(−)
A8	733	(−)	(−)	(+)	≈
Number of RAPD products		17–23	13–25	13–24	9–15
Interpretive group size (n)		8	22	12	6

† Based on cluster analysis using the presence or absence of RAPD bands from each accession.
‡ Indicates RAPD product used in stepwise discriminant analysis function to group accessions into genetic diversity pools.
§ +, −, (+), (−), and ≈ indicate the presence, absence, mostly present (>60%), mostly absent (<40%), and general equally absent and present bands (40–60%), respectively, of the polypeptide class indicated for each interpretive group.
¶ Indicates RAPD product used to place supplemental accessions into one of four genetic diversity pools.

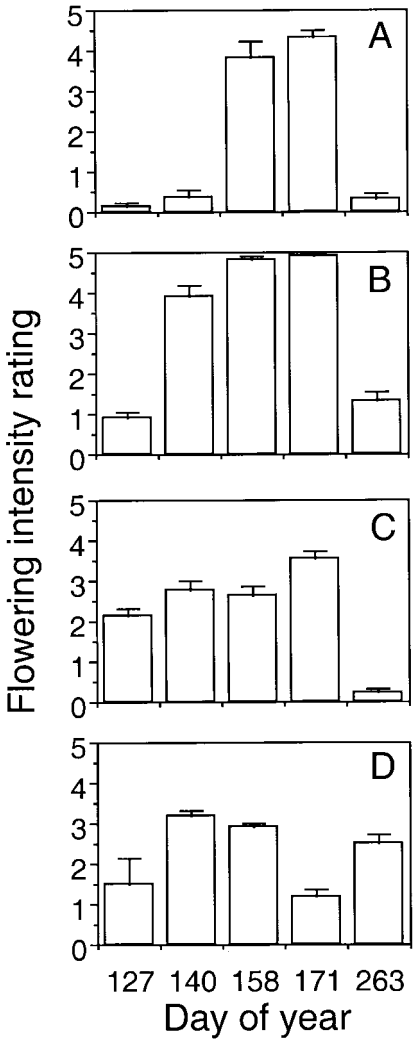

Fig. 3. The four flowering intensity rating classes used to classify the 48-accession birdsfoot trefoil core subset collection. The sub-figures A, B, C, and D correspond to Table 1 interpretive groups 1, 2, 3, and 4, respectively. Flowering intensity is a score of 1 to 5: 1, not flowering; 2, very few flowers; 3, few to moderate flowering; 4, moderate; and 5, intense flowering.

portion of the total variation. In our study, the amount of variation accounted for varied depending upon the descriptor. For SGPP, RAPD, flowering pattern, and FRI, the first three PCs accounted for 73, 38, 92, and 100% of the combined total variation, respectively. In particular, the RAPD data were not robustly represented by the first three principal components, so significant amounts of potentially useful classification information were lost. Other recent publications have reported similar ranges of total variation accounted for by the first few principle components using similar kinds descriptors as in this study: alfalfa (*Medicago sativa* L.; several morphologic traits, 3-PCs, 60%; Smith et al., 1995); barley (*Hordeum vulgare* L.; RFLPs, 2-PCs, 41%; Melchinger et al., 1994); coconut palm (*Cocos nucifera* L.; RAPDs, 3-PCs, 35%; Ashburner et al., 1997); cotton (*Gossypium arboreum* L. and *G. herbaceum* L.; morphological traits, 3-PCs, 30 to 41%; Stanton et al., 1994); open-pollinated corn (*Zea mays* L.; isozymes, 2-PCs, 49%, Revilla and Tracy; 1995); ground nut (*Vigna subterranea* L.; isozymes, 2-PCs, 83–92%; Pasquet et al., 1999); perennial ryegrass (*Lolium perenne* L.; morphological traits, 1-PC, 46–67%; Casler, 1995); rice (*Oryza sativa* L.; seed and agronomic traits, 2-PCs, 65%; Mackill and Lei, 1997); wild potato species (*Solanum* sect. *Petota*; RFLPs, 2-PCs, 71% and morphology, 2-PCs, 70%; Clausen and Spooner, 1998); soybean [*Glycine max* (L.) Merr.; RFLPs, 23%, Diers et al., 1997]; and triticale (×*Triticosecale* Wittmack; morphological and cytological traits, 4-PCs, 82%; Furman et al., 1997). The small percentage of total variation accounted for by principal components analysis from some of these other reports suggests the interpretive group approach may be a valuable alternative for interpreting collections of other species.

Table 5. Means and differences for seven ecogeographic descriptors from four genetic diversity pools using analysis of variance to describe the 48 accession birdsfoot trefoil core subset from the USDA-ARS National Plant Germplasm System collection.

Genetic diversity pool†	*n*	Ecogeographic descriptor‡							
		Low temperature§	High temperature	Sunshine	Precipitation	Snow	Adjusted latitude	Longitude	Elevation
		°C		%	mm		°		
Asia Minor	8	−2.1b	20.7ab	62.0a	72.2	26.8	37.9bc	38.9a	1.0a
European	26	0.6b	18.9b	42.9b	78.9	19.5	46.7a	−10.9b	0.3b
Mediterranean	5	7.6a	23.7a	60.2a	57.7	3.1	32.7c	22.4ab	0.4b
North American	9	−7.1c	22.1a	53.8a	76.2	40.2	44.0ab	−76.7c	0.2b
Significant difference		***	**	***	ns	ns	***	***	**

* Indicates significance at $P \leq 0.05$ on the basis of Fisher's least significance test (LSD). Means within columns followed by the same letter are not significantly different according to LSD.
** Indicates significance at $P \leq 0.01$ on the basis of Fisher's least significance test (LSD). Means within columns followed by the same letter are not significantly different according to LSD.
*** Indicates significance at $P \leq 0.001$ on the basis of Fisher's least significance test (LSD). Means within columns followed by the same letter are not significantly different according to LSD.
ns = nonsignificant.
† Genetic diversity pools were determined by a maximum parsimony analysis using teh multistate classes of the eight plant interpretive descriptor groups.
‡ Plant descriptors: Adjusted latitude, absolute value of latitude to compensate for collecting sites in South America; Elevation, collecting site origin classified as either lowland or highland origins, 0 and 1, respectively, based on Bailey (1989).

Genetic Diversity Pools

The structure of accessions in the core subset was defined with the nine interpretive group variables using heuristic analysis (Table 1). The 48 accessions were classified into four genetic diversity pools (Fig. 2) and stepwise discriminant analysis verified accession placement (100% correctly classified). The significant variables in the classification were three SGPP (Table 3) and six RAPD (Table 4) markers, herbage tannin content, and flowering intensity at the second field sampling time (Day of Year 140). The canonical correlation coefficients were 0.98, 0.93, and 0.89, with eigenvalues of 28.3, 6.6, and 3.8, respectively. The descriptor characteristics for the accessions in each genetic diversity pool were also interpreted by the ecogeographic origins of the accessions (Table 5). The low and high temperatures and yearly sunshine hours were the ecogeographic variables that differentiated the four interpretive groups (Table 5). Annual precipitation and snow amounts had no distinguishing effects.

The Asia Minor pool was comprised of eight wild accessions collected primarily in Iran and Turkey (Table 1). All of the accessions were collected in upland ecoregions (Table 5), with many exhibiting atypical morphology that included either pilose or pubescent leaves and some plants displaying a prostrate growth habit (Steiner and Poklemba, 1994). All of the accessions in the Asia Minor pool were later flowering compared with the other three pools, but maintained intense flowering throughout the summer growing season (Fig. 3C) and generally had the greatest FRI of the four pools. The greater the FRI, the more rapidly a plant population flowers when exposed to a range of photoperiod lengths. These accessions were also among the lowest for herbage tannin content, pod weight, leaf size, and seed chalcid infestation for all pools (Table 6). Since these accessions were from mountainous collecting sites, high FRI and intense flowering could contribute to increased seed production under limited flowering period conditions (Steiner and Garcia de los Santos, 2001). The Asia minor accessions were variable for SGPP and RAPD markers (similarity indices equal 0.43 and 0.46, respectively).

The 26 accessions in the European genetic diversity pool were collected across Europe and included the European cultivars Late Roskilde II and Stirpe 13; U.S. cultivars AU Dewey, Kalo, NC-83, and Viking; and four naturalized populations collected in Canada and Uruguay (Table 1). European pool accessions generally exhibited earlier flowering than accessions from the other three pools (Table 6), but flowered less intensely throughout the summer (Fig. 3C), and had the lowest FRI when grown under a 13-h photoperiod length. The European accessions had greater herbage tannin contents compared to accessions in the Asia Minor and North American pools, and were the most susceptible to seed chalcid infestation (Table 6). These accessions were generally placed in SGPP interpretive Groups 1 and 3 (64 and 28%, respectively) and 88% in RAPD Group-2. Two-thirds of these accessions were lowland ecogeographic types (Steiner and Poklemba, 1994). The U.S. cultivars in the European pool had pedigrees with parental sources tracing directly back to European accessions (Steiner and Poklemba, 1994), and were generally different from most other North American cultivars. Similarly, the naturalized populations from Canada and Uruguay were different from the North American pool cultivars.

The five Mediterranean pool accessions were geographically diverse, ranging from Asia Minor to the western Mediterranean region and including Ethiopia. The accessions in this pool generally had atypical birdsfoot trefoil morphology compared to accessions in the other pools, having the largest leaves and greatest tannin content (Table 6). They were the only accessions classified in RAPD Group-4. A single accession in the core subset from Ethiopia (PI 273937) represented a group of naturally occurring autogamous tetraploid genotypes found in the NPGS collection (Steiner and Poklemba, 1994). The French (PI 235525) and Spanish (PI 246735) accessions were morphologically similar to the Ethiopian accession with thick stems, but were not autogamous. Additionally, all accessions in this pool generally had pilose or pubescent leaves and stems that grew prostrate (Steiner and Poklemba, 1994).

Table 6. Means and differences for nine plant descriptors from four genetic diversity pools using analysis of variance to describe the 48 accession birdsfoot trefoil core subset from the USDA-ARS National Plant Germplasm System collection.

Genetic diversity pool†	*n*	Plant descriptor‡								
		Leaf size§	Pod weight	Chalcid	Tannin	Early flowering	FRI	FRI-13	FRI-16	FRI-19
		mm^2	g 10^2	$umbel^{-1}$	m eq. g^{-1}	%				
Asia Minor	8	26.1b	4.9b	3.7b	5.1c	10.8b	4.0	0.29a	1.75	1.99
European	26	42.5b	6.7a	5.7a	24.1b	25.8a	3.8	0.04b	1.69	2.07
Mediterranean	5	100.1a	5.8ab	3.5b	57.8a	6.3b	4.2	0.16ab	1.86	2.20
North American	9	43.2	7.2a	4.5ab	4.3c	1.9b	3.5	0.05b	1.55	1.93
Significant difference		***	**	*	***	***	ns	*	ns	ns

* Indicates significance at $P \leq 0.05$ on the basis of Fisher's least significance test (LSD). Means within columns followed by the same letter are not significantly different according to LSD.
** Indicates significance at $P \leq 0.01$ on the basis of Fisher's least significance test (LSD). Means within columns followed by the same letter are not significantly different according to LSD.
*** Indicates significance at $P \leq 0.001$ on the basis of Fisher's least significance test (LSD). Means within columns followed by the same letter are not significantly different according to LSD.
ns = nonsignificant.
† Genetic diversity pools were determined by a maximum parsimony analysis using the multistate classes of the eight plant interpretive descriptor groups.
‡ Plant descriptors: Leaf, size of center leaflet (length × width); Pod, weight of an individual dried umbel of pods; Chalcid, number of seed chalcids per umbel; Tannin, herbage tannin content; Flower, flowering pattern for five sampling times; Early flowering, percentage of clones that had flowered on 7 May; FRI, aggregate flowering response index; FRI-13, FRI at 13 hr photolength; FRI-16, FRI at 16 hr photolength; FRI-19 at 19 hr photolength.

The nine North American genetic diversity pool accessions were primarily comprised of cultivars selected in the USA and Canada, most of which had the cultivar Empire or other common progenitors in their pedigrees (Steiner and Poklemba, 1994). Only one wild accession (PI 251143) from the Balkan region in Europe was placed in the North American pool. The accessions in this pool had the lowest early flowering percentage and flowered the least under 13-h photoperiod lengths, compared with accessions in all other pools (Table 6). North American accessions were generally had low herbage tannin content, uniform SGPP marker distributions (I = 0.81), were among the most susceptible to seed chalcid infestation and had the greatest seed pod size (Table 6). All of the North American accessions, as well as most accessions in the Asia Minor and Mediterranean genetic diversity pools, had the lowest percentages for early flowering (Table 2). It is interesting that the North American accessions were generally the only ones placed in RAPD Group-1 (Table 1). Using RAPD profiles as an indicator of genetic background, this finding suggests that plant breeding selection pressures used to develop North American cultivars have generally shifted these materials from their wild or naturalized progenitors found in the Old World. The North American accessions were placed between the very divergent European and Asia Minor/Mediterranean accessions in the cladistic analysis tree (Fig. 2), further indicating their uniqueness from Old World materials. An earlier study of birdsfoot trefoil accessions from the NPGS active collection also grouped North American accessions apart from Old World materials (McGraw et al., 1989).

Plant and Ecogeographic Descriptor Relationships

Extensive compilations of multiple traits have not been reported for NPGS core subset collections. Doing so should help to understand relationships among different traits (Brown, 1989). For the birdsfoot trefoil core subset, we found the only simple correlations among plant descriptors were herbage tannin content with leaf size (r = 0.69; $P \leq 0.001$), and 16 with 19-h photoperiod FRI (r = 0.0.74; $P \leq 0.001$). This indicated different plant traits contributed unique information to the overall accession classification structure, but without introducing a bias because of collinearity among descriptors. Similarly, simple descriptors such as pod size, chalcid infestation, and leaf size were usually not associated with multistate plant characteristics such as SGPP or RAPD (Table 7), so simple and multistate characters uniquely contributed to the classification. Only herbage tannin content was associated with SGPP, RAPD (SGPP and RAPD were collinear), and flowering pattern. These results suggest selection for specific traits can be made from materials with multiple genetic backgrounds. Thus genetic bottlenecks resulting from repeated selections from the same sources could be avoided as has been a common problem with many forage cultivars (Rumbaugh, 1991), including birdsfoot trefoil (Steiner and Poklemba, 1994).

Table 7. Product moment correlation (r_{pm}) for comparing three multi-character descriptors among one another and with five simple plant descriptors for the 48-accession NPGS birdsfoot trefoil core subset collection.

Plant descriptor†	SGPP	Flowering pattern	FRI	Early flowering	Tannin	Pod	Chalcid	Leaf size
	r_{pm}							
RAPD	0.30**	0.33**	0.20*	0.11	0.35**	−0.02	0.16*	0.09
SGPP		0.29**	0.09	−0.03	0.72**	−0.03	0.10	0.15
Flowering			0.21**	0.39**	0.37**	−0.08	0.11	0.22**

* Indicates significance at $P \leq 0.05$.
** Indicates significance at $P \leq 0.01$.
† RAPD, genetic distance based on random amplified polymorphic DNA products; SGPP, water soluble seed globulin polypeptides; and Flowering, flowering pattern.

Examining the relationships of traits compiled from NPGS core subset evaluations with ecogeographic features of accession origins could provide curators and users better ways to manage and use germplasm. The only correlation between a simple plant descriptor with an ecogeographic descriptor was herbage tannin content with low temperature ($r = 0.66$; $P \leq 0.001$). Similarly, for multicharacter descriptors, SGPP was associated with collecting site low and high temperature and latitude (Table 8). The RAPD descriptor was related to most ecogeographic features, and flowering pattern similarities were associated with collecting site low temperature, latitude, and longitude. Also, the geographic distances between collecting sites was correlated with genetic distance (Table 8).

Previous birdsfoot trefoil research showed that morphological traits from wild genotypes were associated with ecogeographic features of collecting site habitats (Steiner, 1999; Steiner and Garcia de los Santos, 2001). These studies utilized only wild or naturalized materials, while our study included cultivars. Similarly, McGraw et al. (1989) also included cultivars in their study and did not find correlations between the variables they measured and the geographic origins of the accessions. We found that when cultivars are included in ecogeographic analyses, the effect is random and so their exclusion from the analysis increases the correlation coefficient (herbage tannin content with low temperature, $r = 0.72$; $P \leq 0.004$), without changing which relationships are significant or not. It is possible that core subset collections that were selected on the basis of country of origin, as with the birdsfoot trefoil collection, may reveal few relationships between accessions and their ecogeographic origins. Further research is needed to determine whether or not relationships exist in other collections.

Supplementary Accession Placement

A problem in ex situ germplasm collection management is how to deal with the addition of newly acquired accessions to the active collections after the core subset has been identified. Often it is not known whether the newly acquired accessions are similar to or different than existing accessions in the collection. To address this problem, we proposed using the core subset as a reference collection to describe the range of variation in the active collection (Beuselinck and Steiner, 1992).

On the basis of this principle, three accessions not included in the original core subset were evaluated for their relative similarity to the 48-subset accessions (Table 1).

Description of Supplementary Accessions

The three accessions chosen were also tetraploids and were selected for apparent morphologic differences from those accessions already in the core subset (Steiner and Poklemba, 1994). Both G 31276 and PI 234670 exhibited rhizomatous root growth that was not previously described in *L. corniculatus*. G 31276 was collected in Morocco and recently entered into the GRIN system (Beuselinck et al., 1996). The derivation of the rhizome character in *L. corniculatus* is not known, but G 31276 may have acquired the trait from the rhizominous species *L. uliginosus* Schkuhr. (Steiner, 1999). Even though both AU Dewey (Pedersen et al., 1986) and Kalo (D. Darris,1996, personal communication) in the core subset are reported to have rhizomes, neither exhibited the trait in this experiment but were also placed in the European pool. PI 234670 was entered into GRIN in 1956, but was not reported to display rhizomes until recently (Steiner and Garcia de los Santos, 2001). PI 234811 displayed leaf morphology atypical of most *L. corniculatus* (Steiner and Poklemba, 1994), was extremely self-incompatible (Garcia de los Santos et al., 2001), and represented a family of accessions from alpine environments in the Swiss Alps. Accessions from this region are considered by some as *L. alpinus* (DC.) Schleich. ex Raymond (Zertova, 1964) and to be progenitors of *L. corniculatus* (Reynaud and Jay, 1991). This accession also exhibited an adventitious rhizome production in response to wounding, such as occurs when making vegetative plant cuttings (Fjellstrom and Steiner, 1995, unpublished data).

Reference Collection Comparisons

Stepwise discriminant analysis classified the three supplementary accessions using six RAPD markers, herbage tannin content, and 13-h photoperiod FRI. The canonical correlation coefficients were 0.96, 0.89, and 0.76 with eigenvalues of 10.4, 3.9, and 1.4, respectively. Accessions G 31276, PI 234670, and PI 234811 were placed into the European, European, and Asia Minor genetic diversity pools, respectively (Table 1). By means of a reduced plant descriptor data set (pod, chalcid, early flowering, and flowering pattern data were not available for the supplementary accessions), the stepwise discriminant analysis correctly replaced 47 of the 48 accessions in the core subset (98%) into the four genetic diversity pools (Table 1). Iranian PI 228286 was placed in the Mediterranean instead of Asia Minor pool. Reassigning the misplaced accession did not change the

Table 8. Product moment correlation (r_{pm}) for comparing three multi-character descriptors with eight quantitative ecogeographic descriptors for the 48-accession NPGS birdsfoot trefoil core subset collection.

Plant descriptor†	Low temperature	High temperature	Sunshine	Precipitation	Snow	Adjusted latitude	Longitude	Geographic distance
	r_{pm}							
RAPD	0.17*	0.10	0.17*	0.12*	0.07	0.14*	0.15*	0.11*
SGPP	0.27*	0.12*	−0.01	0.08	−0.08	0.31*	−0.08	0.05
Flowering	0.22*	0.09	0.19*	0.08	0.02	0.25*	0.18*	0.17*

* Indicates significance at $P \leq 0.05$.

† RAPD, genetic distance based on random amplified polymorphic DNA products; SGPP, water soluble seed globulin polypeptides; and Flowering, flowering pattern.

genetic diversity pool placement of the three supplementary accessions.

The three supplementary accessions provided novel trait diversity, complimentary to the existing core subset accessions. The presence of rhizomes in G 31276 and PI 234670 would make these accessions likely candidates for inclusion into the core subset. Also, PI 234811 is geographically distinct from the other Asia Minor accessions (most geographically western collected accession in this genetic diversity pool). All three accessions had interpretive group profiles different than all accessions in their respective genetic diversity pools.

Collection management decisions need to be considered regarding what to do when unique accessions are identified. One option would be to consider the three for inclusion in the core subset. A second option would be to replace unremarkable or duplicative accessions with the supplemental accessions. Since G 31276 and PI 234670 were placed in the European genetic diversity pool, they could replace either PI 233807 or PI 262530 which have the same interpretive group profiles as the accession preceding them (Table 1). The European pool has the greatest number of accessions of the four genetic diversity pools, so substitution seems to be the best option. In the case of PI 234811, the Asia Minor genetic diversity pool is relatively small, so addition may be a reasonable option. A third option could be to not change the core subset, but make note of unique accessions. The use of core subset plant descriptor data as a reference collection in conjunction with discriminant analysis appears to be a suitable method for classifying accessions not included in core subset collections.

CONCLUSION

In this research, the concepts of interpretive groups, genetic diversity pools, and reference collection comparisons for assessing and managing core subset collections were presented and applied to the NPGS birdsfoot trefoil core subset collection. Interpretive groups derived from categorized simple qualitative or quantitative trait or multistate descriptors described accession variability provided and allowed a way for trait comparisons on a common scale. By means of cladistic analysis of the interpretive group descriptions, four distinct genetic diversity pools were identified that differed by ecogeographic origins. The accessions from North American were comprised mainly of cultivars different from those originating in the Old World and classified as Asia Minor, European, and Mediterranean. Accessions from the Old World genetic diversity pools contained traits similar to most accessions found in the North American pool, but RAPD analyses showed their genetic backgrounds to be different. This suggests that new cultivars could be developed from genetic sources different from those used to develop the North American cultivars. Additionally, the core subset collection was shown to be useful as a reference collection to assess the uniqueness of three accessions acquired after establishment of the core subset. These approaches should be useful for making decisions whether to make future acquisitions a part of active collections, or their inclusion in core subset collections. The concepts of interpretive groups, genetic diversity pools, and reference collection comparisons should be applicable for assessing and managing other core subset collections.

ACKNOWLEDGMENTS

Thanks to the technical assistance of C.J. Poklemba in the laboratory and W.E. Gavin in the field. This research was funded in part by grants from the Clover and Special Purpose Legume Crop Germplasm Committee.

REFERENCES

Ashburner, G.R., W.K. Thompson, and G.M. Halloran. 1997. RAPD analysis of south Pacific coconut palm populations. Crop Sci. 37: 992–997.

Avise, J.C. 1994. Molecular markers, natural history, and evolution. Chapman & Hall, New York.

Bailey, R.G. 1989. Explanatory supplement to ecoregions map of the continents. Environ. Conserv. 16:307–310.

Beuselinck, P.R., B. Li, and J.J. Steiner. 1996. Rhizomatous *Lotus corniculatus* L.: I. Taxonomic and cytological study. Crop Sci. 36:179–185.

Beuselinck, P.R., and J.J. Steiner. 1992. A proposed framework for identifying, quantifying, and utilizing plant germplasm resources. Field Crops Res. 29:261–272.

Brown, A.H.D. 1989. Core collections: a practical approach to genetic resources management. Genome 31:818–824.

Casler, M.D. 1995. Patterns of variation in a collection of perennial ryegrass accessions. Crop Sci. 35:1169–1177.

Chang, A.T.C., J.L. Foster, D.K. Hall, H.W. Powell, and Y.L. Chien. 1990. Nibus-7 SMMR derived global monthly snow cover and snow depth data set for October 1978 through August 1987. *In* Global ecosystems database, Version 1.0: disc B, documentation manual. Key to geophysical records documentation no. 27. USDOC/NOAA National Geophysical Data Center, Boulder, CO.

Clausen, A.M., and D.M. Spooner. 1998. Molecular support for the hybrid origin of the wild potato species *Solanum* × *rechei*. Crop Sci. 38:858–865.

Diers, B.W., H.T. Skorupska, A.P. Rao-Arelli, and S.R. Cianzio. 1997. Genetic relationships among soybean plant introductions with resistance to soybean cyst nematodes. Crop Sci. 37:1966–1972.

English, J.T., and P.R. Beuselinck. 2000. Methods for evaluating birdsfoot trefoil for susceptibility to foliar and shoot blight caused by *Rhizoctonia* spp. Crop Sci. 40:841–843.

Frankel, O.H. 1984. Genetic perspectives of germplasm conservation. p. 161–170. *In* W. Arber et al. (ed.) Genetic manipulation: Impact on man and society. Cambridge University Press, Cambridge, UK.

Furman, B.J., C.O. Qualset, B. Skovmand, J.H. Heaton, H. Corke, and D.M. Wesenberg. 1997. Characterization and analysis of North American triticale genetic resources. Crop Sci. 37:1951–1959.

Garcia de los Santos, G., J.J. Steiner, and P.R. Beuselinck. 2001. Adaptive ecology of *Lotus corniculatus* L. genotypes: II. Crossing ability. Crop Sci. 41:564–570.

Goodman, M.M. 1972. Distance analysis in biology. Syst. Zool. 21:174–186.

Greene, S.L., and J.R. McFerson. 1996. Conservation of *Lotus* genetic resources: Status of the U.S. collection. p. 39–42. *In* P.R. Beuselinck and C.A. Roberts (ed.) First International *Lotus* Symposium, 22–24 March 1994, St. Louis, MO. Univ. Exten., Univ. of Missouri, Columbia.

Greene, S.L., T.C. Hart, and A. Afonin. 1999a. Using geographic information to acquire wild crop germplasm for ex situ collections: I. Map development and field use. Crop Sci. 39:836–842.

Greene, S.L., T.C. Hart, and A. Afonin. 1999b. Using geographic information to acquire wild crop germplasm for ex situ collections: II. Post-collection analysis. Crop Sci. 39:843–849.

Kimura, M. 1982. Molecular evolution, protein polymorphism and evolutionary theory. Japan Scientific Societies Press, Tokyo.

Kinemann, J.J., and M.A. Ohrenschall. 1992. Global ecosystems data base, Version 1.0: disc A, documentation manual. Key to geophysi-

cal records documentation no. 27. USDOC/NOAA National Geophysical Data Center, Boulder, CO.
Kinemann, J.J., and M.A. Ohrenschall. 1994. Global ecosystems database, Version 1.0: disc B, beta test. USDOC/NOAA National Geophysical Data Center, Boulder, CO.
Larsen, K. 1954. Cytotaxonomical studies in *Lotus*. I. *Lotus corniculatus* L. sens. lat. Bot. Tidsskr. 51:205–211.
Leemans, R., and W.P. Cramer. 1992. The IIASA database for mean monthly values of temperature, precipitation, and cloudiness on a global terrestrial grid. *In* Global ecosystems database, Version 1.0: disc A, documentation manual. Key to geophysical records documentation no. 27. USDOC/NOAA National Geophysical Data Center, Boulder, CO.
Mackill, D.J., and X. Lei. 1997. Genetic variation for traits related to temperate adaptation of rice cultivars. Crop Sci. 37:1340–1346.
Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27:209–220.
Marshall, D.R. 1989. Limitations to the use of germplasm collections. p. 105–120. *In* A.H.D. Brown, O.H. Frankel, D.R. Marshall, and J.T. Williams (ed.) The use of plant genetic resources. Cambridge University Press, Cambridge, UK.
McGraw, R.L., P.R. Beuselinck, and G.C. Marten. 1989. Agronomic and forage quality attributes of diverse entries of birdsfoot trefoil. Crop Sci. 29:1160–1164.
Melchinger, A.E., A. Graner, M. Singh, and M.M. Messner. 1994. Relationship among European barley germplasm: I. Genetic diversity among winter and spring cultivars revealed by RFLPs. Crop Sci. 34:1191–1199.
Pasquet, R.S., S. Schwedes, and P. Gepts. 1999. Isozyme diversity in bambara groundnut. Crop Sci. 39:1228–1236.
Pedersen, J.F., R.L. Haaland, and C.S. Hoveland. 1986. Registration of 'AU Dewey' birdsfoot trefoil. Crop Sci. 26:1081.
Revilla, P., and W.F. Tracy. 1995. Isozyme variation and phylogenetic relationships among open-pollinated sweet corn cultivars. Crop Sci. 35:219–227.
Reynaud, J., and M. Jay. 1991. Evolution and differentiation of *Lotus corniculatus*/*Lotus alpinus* populations from the French southwestern Alps. II. Contribution of phenolic metabolism markers. Evol. Trends Plants 5:149–155.
Rieseberg, L.H., and S.J. Brunsfeld. 1992. Molecular evidence and plant introgression. p. 151–176. *In* P.S. Soltis et al. (ed.) Molecular systematics of plants. Chapman and Hall, New York.
Roberts, C.A., P.R. Beuselinck, M.R. Ellerieck, D.K. Davis, and R.L. McGraw. 1993. Quantification of tannins in birdsfoot trefoil germplasm. Crop Sci. 33:675–679.
Rohlf, F.J. 1997. NTSYS-pc. Numerical taxonomy and multivariate analysis system. Version 2.2 Applied Biostatistics Inc., Setauket, NY.
Rumbaugh, M.D. 1991. Plant introductions: The foundation of North American forage legume cultivar development. p. 103–114. *In* H.L. Shands and L.E. Wiesner (ed.) Use of plant introductions in cultivar development. CSSA Special Publ. 17. ASA, CSSA, and SSA, Madison, WI.
Small, E., W.F. Grant, and C.W. Crompton. 1984. A taxonomic study of the *Lotus corniculatus* complex in Turkey. Can J. Bot. 62: 1044–1053.
Smith, S.E., L. Guarino, A. Al-Doss, and D.M. Conta. 1995. Morphological and agronomic affinities among Middle Eastern alfalfas—accessions from Oman and Yemen. Crop Sci. 35:1188–1194.
Snedecor, G.W., and W.G. Cochran. 1980. Statistical methods. Iowa State Univ. Press, Ames, IA.
Spagnoletti Zeuli, P.L., and C.O. Qualset. 1987. Geographical diversity for quantitative spike characters in a world collection of durum wheat. 1987. Crop Sci. 27:235–241.
Stanton, M.A., J. McD. Stewart, A.E. Percival, and J.F. Wendel. 1994. Morphological diversity and relationships in the A-genome cottons, *Gossypium arboreum* and *G. herbaceum*. Crop Sci. 34:519–527.
Steiner, J.J. 1999. Birdsfoot trefoil origins and germplasm diversity. p. 81–96. *In* P.R. Beuselinck (ed.) Trefoil: The science and technology of *Lotus*. ASA-CSSA Special Publ. 28. Madison, WI.
Steiner, J.J., and G. Garcia de los Santos. 2001. Adaptive ecology of *Lotus corniculatus* L. genotypes: I. Plant morphologic and RAPD marker characterizations. Crop Sci. 41:552–563.
Steiner, J.J., and S.L. Greene. 1996. Proposed ecological descriptors and their utility for plant germplasm collections. Crop Sci. 36:439–451.
Steiner, J.J., and C.J. Poklemba. 1994. *Lotus corniculatus* classification by seed globulin polypeptides and relationship to accession pedigrees and geographic origin. Crop Sci. 34:255–264.
Swofford, D.L. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods), Version 4. Sinauer Associates, Sunderland, MA.
Warburton, M.L., and S.E. Smith. 1993. Regional diversity in nondormant alfalfas from India and the Middle East. Crop Sci. 33:852–858.
Ward, J.H. 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58:236–244.
Zertova, A. 1964. Bemerkungen zur *Lotus alpinus* Schleich. Ex Ser. In Pesterreich. Oesterr. Bot. Z. 11:337–344.